CREATE TABLE cyclistic(ride\_id VARCHAR, rideable\_type TEXT, start\_date VARCHAR, start\_time VARCHAR, end\_date VARCHAR, end\_time VARCHAR, start\_station\_name VARCHAR, start\_station\_id VARCHAR, end\_station\_name VARCHAR, end\_station\_id VARCHAR, start\_lat VARCHAR, start\_lng VARCHAR, end\_lat VARCHAR, end\_lng VARCHAR, member\_casual TEXT);

SELECT ride\_id, COUNT(\*)

FROM cyclistic

GROUP BY ride\_id

HAVING COUNT(\*) > 1; /\* no duplicates\*/

SELECT \* FROM cyclistic WHERE start\_lat = '41.87' AND start\_lng = '-87.61'; /\*all NA for start station name and id\*/

DELETE FROM cyclistic WHERE start\_station\_name = 'NA' AND start\_station\_id = 'NA'; /\*DELETE 113410\*/

DELETE FROM cyclistic WHERE end\_station\_name = 'NA' AND end\_station\_id = 'NA'; /\*DELETE 67448\*/

SELECT \* FROM cyclistic; /\*867717\*/

SELECT DISTINCT member\_casual FROM cyclistic; /\* member, casual\*/

SELECT \*

FROM cyclistic

WHERE start\_station\_name = 'NA' OR end\_station\_name = 'NA' OR

(start\_station\_id = 'NA' OR end\_station\_id = 'NA');

/\*none\*/

ALTER TABLE cyclistic

ADD ride\_time INT;

SELECT DISTINCT start\_station\_name FROM cyclistic; /\* 1182\*/

UPDATE cyclistic

SET ride\_time = (DATE\_PART('day', CONCAT(end\_date, ' ', end\_time)::timestamp - CONCAT(start\_date, ' ', start\_time)::timestamp) \* 24 +

               DATE\_PART('hour', CONCAT(end\_date, ' ', end\_time)::timestamp - CONCAT(start\_date, ' ', start\_time)::timestamp)) \* 60 +

               DATE\_PART('minute', CONCAT(end\_date, ' ', end\_time)::timestamp - CONCAT(start\_date, ' ', start\_time)::timestamp); /\* updated 867717\*/

SELECT \*

FROM public.cyclistic

WHERE start\_date = '5/23/2022' and end\_date = '5/24/2022';

/\*to check if ride\_time is valid\*/

select distinct ride\_time from (SELECT ride\_time

FROM cyclistic

WHERE ride\_time <=0) as dis\_ride\_time; /\* -129, -127,0\*/

SELECT \*

FROM cyclistic

WHERE ride\_time = -129 or ride\_time = -127; /\* 2 rows with wrong data\*/

DELETE FROM cyclistic WHERE ride\_time = -129 or ride\_time = -127 or ride\_time = 0; /\*delete 8418\*/

select \* from cyclistic where ride\_time<=0; /\*0 = confirmation\*/

/\*Member Top 5\*/

SELECT start\_station\_name, count(\*) as occurrence, AVG(ride\_time)

FROM cyclistic

WHERE member\_casual = 'member'

GROUP BY start\_station\_name

ORDER BY occurrence DESC, AVG(ride\_time) DESC;

Top 5 locations:

"DuSable Lake Shore Dr & North Blvd" 4300 18.5120930232558140

"Clark St & Elm St" 3786 12.4601162176439514

"Kingsbury St & Kinzie St" 3763 10.2017007706617061

"Clinton St & Washington Blvd" 3671 11.4165077635521656

"Clinton St & Madison St" 3530 11.2419263456090652

/\*Member Least 5\*/

SELECT start\_station\_name, count(\*) as occurrence, AVG(ride\_time)

FROM cyclistic

WHERE member\_casual = 'member'

GROUP BY start\_station\_name

ORDER BY occurrence ASC, AVG(ride\_time) ASC;

Top 5 locations for immediate improvement:

"Public Rack - King Dr & Oakwood Blvd" 1 1.00000000000000000000

"Western Ave & 92nd Pl" 1 1.00000000000000000000

"Ashland Ave & 45th St" 1 1.00000000000000000000

"Major Taylor Trail & 124th St" 1 1.00000000000000000000

"Martin Luther King Dr & 111th St" 1 2.0000000000000000

125 locations including above 5 which are not popular can either be discarded and their funds concentrated on other locations or if funds are available it can be used for ads.

/\*Casual Top 5\*/

SELECT start\_station\_name, count(\*) as occurrence, AVG(ride\_time)

FROM cyclistic

WHERE member\_casual = 'casual'

GROUP BY start\_station\_name

ORDER BY occurrence DESC, AVG(ride\_time) DESC;

Top 5 locations:

"Streeter Dr & Grand Ave" 9755 38.0824192721681189

"DuSable Lake Shore Dr & Monroe St" 8552 38.2547942001870907

"Michigan Ave & Oak St" 6856 37.5138564760793466

"DuSable Lake Shore Dr & North Blvd" 6418 29.9111872857588034

"Millennium Park" 5191 41.7692159506838759

/\*Casual Least 5\*/

SELECT start\_station\_name, count(\*) as occurrence, AVG(ride\_time)

FROM cyclistic

WHERE member\_casual = 'casual'

GROUP BY start\_station\_name

ORDER BY occurrence ASC, AVG(ride\_time) ASC;

Top 5 locations for immediate improvement:

"Melvina Ave & Belmont Ave" 1 1.00000000000000000000

"Western Ave & 111th St - NW" 1 1.00000000000000000000

"Public Rack - Chase Ave & Touhy Ave - NE" 1 1.00000000000000000000

"Public Rack - Sacramento Ave & Devon Ave" 1 1.00000000000000000000

"Public Rack - Narragansett Ave & Fullerton Ave" 1 3.0000000000000000

95 locations including above 5 which are not popular can either be discarded and their funds concentrated on other locations or if funds are available it can be used for ads.

DuSable Lake Shore Dr & North Blvd is popular with both members and casual riders and ranks among top 5. This place can be given more priority for improved satisfaction. 26 locations are neither popular with members nor casual riders. They can either be closed to concentrate funds better on other locations or if funds are in surplus, proper ads can be given to make these locations more visible.